



UWS Academic Portal

Immersive virtual reality as a pedagogical tool in education

Hamilton, D.; McKechnie, J.; Edgerton, E.; Wilson, C.

Published in:
Journal of Computers in Education

DOI:
[10.1007/s40692-020-00169-2](https://doi.org/10.1007/s40692-020-00169-2)

E-pub ahead of print: 11/07/2020

Document Version
Publisher's PDF, also known as Version of record

[Link to publication on the UWS Academic Portal](#)

Citation for published version (APA):
Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2020). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8, 1-32. <https://doi.org/10.1007/s40692-020-00169-2>

General rights

Copyright and moral rights for the publications made accessible in the UWS Academic Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact pure@uws.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design

D. Hamilton¹ · J. McKechnie¹ · E. Edgerton¹ · C. Wilson¹

Received: 21 October 2019 / Revised: 16 June 2020 / Accepted: 2 July 2020
© The Author(s) 2020

Abstract The adoption of immersive virtual reality (I-VR) as a pedagogical method in education has challenged the conceptual definition of what constitutes a learning environment. High fidelity graphics and immersive content using head-mounted-displays (HMD) have allowed students to explore complex subjects in a way that traditional teaching methods cannot. Despite this, research focusing on learning outcomes, intervention characteristics, and assessment measures associated with I-VR use has been sparse. To explore this, the current systematic review examined experimental studies published since 2013, where quantitative learning outcomes using HMD based I-VR were compared with less immersive pedagogical methods such as desktop computers and slideshows. A literature search yielded 29 publications that were deemed suitable for inclusion. Included papers were quality assessed using the Medical Education Research Study Quality Instrument (MERSQI). Most studies found a significant advantage of utilising I-VR in education, whilst a smaller number found no significant differences in attainment level regardless of whether I-VR or non-immersive methods were utilised. Only two studies found clear detrimental effects of using I-VR. However, most studies used short interventions, did not examine information retention, and were focused mainly on the teaching of scientific topics such as biology or physics. In addition, the MERSQI

✉ D. Hamilton
david.hamilton@uws.ac.uk

J. McKechnie
jim.mckechnie@uws.ac.uk

E. Edgerton
edward.edgerton@uws.ac.uk

C. Wilson
claire.wilson@uws.ac.uk

¹ Division of Psychology, School of Education & Social Sciences, University of the West of Scotland, Paisley, UK

showed that the methods used to evaluate learning outcomes are often inadequate and this may affect the interpretation of I-VR's utility. The review highlights that a rigorous methodological approach through the identification of appropriate assessment measures, intervention characteristics, and learning outcomes is essential to understanding the potential of I-VR as a pedagogical method.

Keywords Virtual reality · Head mounted displays · Education · Learning · Simulations

Introduction

The increasing financial feasibility of virtual reality (VR) has allowed for educational institutions to incorporate the technology into their teaching. According to research, 96% of universities and 79% of colleges in the UK are now utilising augmented or virtual reality in some capacity (UKAuthority 2019). In addition, the rising power of personal computers and associated hardware has led to a revolution in graphical fidelity, with ever more complex and realistic simulations and virtual worlds (Slater 2018). As Dickey (2005) alludes to, this has both challenged and expanded the very conceptual definition of what is defined as a learning environment. Where once this would have been restricted to classroom teaching or field trips, VR's innate ability to give users a sense of presence and immersion has opened new possibilities in education if implemented appropriately (Häfner et al. 2018).

The use of technology-aided education as a pedagogical method is not a modern phenomenon, and investigations into its utility have been studied for almost half a century. As far back as the 1970s, Ellinger and Frankland (1976) found that the use of early computers to teach economic principles produced comparative learning outcomes with traditional didactic methods such as lectures. However, as Jensen and Konradsen (2018) allude to, it was with the release of the Oculus Rift in 2013 that VR became synonymous with head-mounted-display (HMD) based VR. This had several ramifications. First, HMDs became economically feasible for consumers and educational institutions to acquire *en masse*, due to a significant drop in price (Hodgson et al. 2015). As Olmos et al. (2018) remarks, the economic viability of VR has tackled one of the main entry barriers to adopting the technology. And secondly, academic research into the potential benefits of I-VR in education starts to expand, as well as its applied use in pedagogical settings (Hodgson et al. 2019). One of VR's most important contributions to education is that it has allowed students to repeatedly practice complex and demanding tasks in a safe environment. This is particularly true of procedural tasks such as surgical operations or dental procedures that cannot be carried out for real until a certain level of competency has been achieved (Alaraj et al. 2011; Larsen et al. 2012). Additionally, VR has allowed for students to gain cognitive skills by way of experiential learning, such as exposing them to environments that would be too logistically problematic to visit in reality (Çalışkan 2011). For instance, by using a HMD, Bailenson et al. (2018) were able to expose students to an underwater environment to facilitate learning about climate change. VR has made an important contribution to education in

that it has allowed for students to directly experience environments or situations that are difficult to replicate by using traditional teaching methods such as lectures, slideshows, or 2D videos.

A concise definition of VR's key characteristics is challenging due to the ever-changing nature of the technology. However, Sherman and Craig (2003) proposed that there are a number of constituent elements that must underpin the VR experience, ultimately leading to the life-like perception of the virtual environment. These include the necessity for VR to be immersive, in that the participant's own cognitive faculties produce a sense of being present and involved in the virtual space, often with reduced awareness of what is happening in the real-world around them. Additionally, the virtual space should offer a degree of interactivity, in that the user can manipulate the environment and test variables. This can include interacting with objects, virtual avatars, or even collaborating with other real-life users within the computer-generated space.

Definition of key terms

Due to the multidisciplinary nature of VR research and its pedagogical applications, it is important to define key terms used. VR can broadly be broken down into two main categories: desktop VR (D-VR), and immersive-VR (I-VR). D-VR is typically classified as non-immersive, in that a headset is not used, and the participant will be controlling and manipulating the virtual environment on a computer screen with traditional keyboard and mouse hardware (Lee et al. 2010). On the other hand, I-VR is typically multi-modal in nature by providing a sense of immersion in the environment through 360° visuals by aid of a HMD, auditory stimulation through the use of earphones, and increasingly the proprioception of limbs by way of controllers and tracking (Freina and Ott 2015; Howard-Jones et al. 2015; Murcia-López and Steed 2016). Although there are a range of HMDs on the market, from high-end hardware like the HTC Vive, to viable low-cost options like the Google Cardboard, they all utilise the same core principals of operation (Brown and Green 2016). Typically, a HMD will feature a set of embedded liquid crystal displays (LCD) which will present each eye an image from a slightly different angle. This mimics natural optic function by allowing the wearer to view a stereoscopic image complete with depth perception and a wide field of view. Mobile VR headsets can achieve the same effect using a single display by dividing the screen down the middle and presenting each half to the corresponding eye. Therefore, the current review defines a HMD as a device worn over the head, which provides a stereoscopic computer-generated or 360° video image to the user. This includes tethered (connected to a computer), stand-alone (no computer needed), or mobile VR headsets (mobile/cell phone connected to a HMD).

Previous literature and reviews

There have been a number of systematic reviews that have previously explored the relationship between VR and pedagogical attainment. Lee (1999) reviewed 19 studies from as far back as 1976 and found that 66% of students in simulation groups outperformed those in their respective control groups. However, this review did not

focus exclusively on an educational level or age range, so featured both young kindergarten children, as well as higher education students. As a result, the generalisability of VR's effectiveness as a pedagogical method is difficult to ascertain, with significant differences in age, task difficulty, and applications. Furthermore, all the studies are dated in terms of the technology utilised and feature early D-VR programmes and rudimentary computer simulations. This early technology may be primitive when compared with the high-fidelity graphics and immersive components of contemporary technology. Nevertheless, these early studies do help exemplify that the use of technology in education is not a new concept, and computer-based simulations have long been employed as a way of facilitating learning.

A more recent analysis was undertaken by Merchant et al. (2014), and looked at three specific sub-categories of VR: games, simulations, and virtual worlds. Games give the actor autonomy and freedom to move around the virtual world, testing hypotheses, achieving goals, and eliciting motivation and learning through immersion (Gee 2004). Simulations attempt to recreate a real-world environment that can help facilitate learning by allowing for the testing of variables and resulting outcomes. Finally, virtual worlds can provide an immersive or non-immersive sense of presence in a three-dimensional (3D) world, and the ability to manipulate, interact, or construct objects. Furthermore, virtual worlds can give the opportunity for multiple users to interact with one another within the digital environment (Dickey 2005). The meta-analysis showed that although game-based VR produced the highest learning outcomes, simulations and virtual worlds were also effective at increasing educational attainment. Once again, the limitation of this review is that it did not restrict its analysis to exclusively one domain of education. Although higher education made up the greatest number of studies, research from elementary and middle school were also included in the analysis.

One of the most recent systematic reviews to look exclusively at I-VR through the utilisation of HMDs was carried out by Jensen and Konradsen (2018). In their comprehensive search of existing literature published between 2013 and 2017, the review identified 21 quantitative and qualitative papers that focused on both learning outcomes in I-VR, and subjective attitudes and experiences on the part of the user. The review found limited effectiveness of HMD in the acquisition of cognitive, psychomotor, and affective skills when compared with less immersive technologies. However, Jensen and Konradsen (2018) did highlight the relatively low quality of studies included as a concern, and this may impede the ability to draw firm conclusions about the educational utility of I-VR.

Rationale for review

There are several fundamental reasons that necessitate an updated assessment of the topic area, such as the increase in relevant published literature, as well as the narrow scope of previous reviews. The last major review looking at I-VR and HMDs as an educational tool was carried out by Jensen and Konradsen (2018), with the most recent studies featured in that paper being published in 2016. Since then, there has been a significant increase in relevant published literature, with > 70% of the

papers included in the current review being published since 2017, and therefore not included in the previous systematic review. Additionally, unlike previous reviews, the current examination of I-VR's pedagogical utility focuses exclusively on studies where I-VR is directly compared to a less immersive method of learning. As a result, the current paper is able to highlight not only whether I-VR is an effective medium, but also whether it is more effective when compared to alternative methods. Additionally, no other systematic review looking at I-VR and HMDs has had a particular focus on the experimental design, assessment measures, and intervention characteristics of the included studies. The review also addresses the underlying methodology of the included studies, to offer an understanding of how I-VR is being employed in experimental literature. Based upon the findings of previous studies as well as areas yet to be sufficiently explored, this paper has a number of core research questions:

- To assess the subject area, discipline, and learning domain that I-VR has been employed in.
- Understand where I-VR confers an educational benefit in terms of quantitative learning outcomes over non-immersive and traditional teaching methods.
- To examine the experimental design of studies, focusing on how learning outcomes are assessed, and how the I-VR intervention is delivered.
- To inform future experimental and applied practice in the field of pedagogical I-VR application.

Methodology

Search strategy

The current systematic review included peer-reviewed journal articles and conference proceedings that passed all the inclusion criteria detailed. An initial scoping review identified seven databases that could be utilised in a comprehensive literature review, as well as associated keywords and search terms. These included Web of Science (Core Collection), Science Direct, Sage, IEEE Xplore, EBSCO, Taylor & Francis, and Google Scholar. These databases encompass a mixture of general, social science, and technological literature.

Each of the seven databases was searched using a series of keywords based on the following Boolean logic string:

("Virtual Reality" OR "Virtual-Reality" OR "Immersive Virtual Reality" OR "Head Mounted Display" OR "Immersive Simulation") AND (Education OR Training OR Learning OR Teaching)

Due to the scope and parameters of the research objectives, only peer-reviewed literature published between January 2013 and December 2018 was included in the final review. Early access articles due to be published in 2019 were also included if these were found using the database searches. Date criteria was based upon an initial scoping review that found a substantial growth in relevant I-VR literature from 2013 onwards. A major contributing factor was the release of the Oculus Rift

Development Kit 1 (DK-1) in early 2013, which is regarded as one of the first economically viable and high quality HMDs that could be used both within educational institutions, and at home (Lyne 2013).

The literature search across the databases yielded more than 12,000 references from a variety of sources. After the removal of duplicate records, 9,359 unique references were included for the title and abstract screening stage of the review.

Selection and screening

The open and general nature of the search string used led to a large number of references being returned for screening. As Jensen and Konradsen (2018) already alluded to in the last major review, VR research transcends various academic disciplines. The result is a lack of a clear taxonomy of definitions and terms. This means a wide net must be cast to ensure comprehensive capture of relevant material. This review defined I-VR as either a completely computer-generated environment, or the viewing of captured 360° video through the use of a HMD. Studies that utilised surgical or dental simulators and trainers such as the da Vinci Surgical System, were excluded as these represent a separate domain of both technological and pedagogical application. For example, surgical simulation based VR typically combines computer-generated visuals with simulated surgical tools, haptic feedback, and robotic components (Li et al. 2017). This type of technology would therefore not be applicable for general pedagogical application. Additionally, references were excluded if they: (1) focused on using I-VR as a rehabilitation or therapeutic tool; (2) were not in English; or (3) where the full-text was not available.

After title and abstract screening was performed, 197 references remained to be included in the full-text review. Each reference had to pass an inclusion flowchart based on each of the following criteria:

1. The population being sampled was from a high school, further or higher education establishment, or was an adult education student.
2. Population sampled did not have a developmental or neurological condition, nor could VR be used as a rehabilitation tool.
3. Paper described an experimental or quasi-experimental trial with at least one control group.
4. At least one group had to have undergone an educational HMD I-VR experience, and was compared with another group who underwent a non-immersive or traditional pedagogical method of education (e.g. Desktop VR, PowerPoint, traditional lecture).
5. A quantitative and objective learning outcome such as tests scores, completion time, or knowledge retention was used to assess effectiveness.

After full-text screening, 29 references passed all stages and were included in the systematic review. See Fig. 1 for a summary of the selection process by stage.

Inter-rater reliability checks were conducted at the title and abstract screening stage to assess the agreement of included studies. There were four individual

evaluators that assessed the suitability of each reference based upon the inclusion criteria, which yielded an average agreement of 96%. Where any disagreement existed, the paper was discussed among all assessors until a unanimous decision was reached as to its suitability.

Quality assessment tool

To assess the quality of the studies, the *Medical Education Research Study Quality Instrument* (MERSQI) was used (Reed et al. 2007). Although this tool was primarily designed to examine the quality of studies in the field of medical education, it is in practice subject neutral. As the MERSQI assesses not only the quality of experimental design and outcomes measures, but also the assessment instrumentation used, it was viewed as a suitable and comprehensive tool for quality appraisal. In addition, the same instrument was used in a previously peer-reviewed systematic review examining VR, by Jensen and Konradsen (2018).

The MERSQI tool covers six quality assessment domains. These include: study design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes. Each domain is scored out of three, with a maximum overall score of 18. Unlike Jensen and Konradsen (2018), the current review gave full points in the study design category for experimental trials with participant randomisation, as well as appropriate pre-intervention measures. This decision was made as true randomised control trials featuring random sampling is unrealistic in I-VR pedagogical research, as the participant sample can only be drawn from an educational establishment.

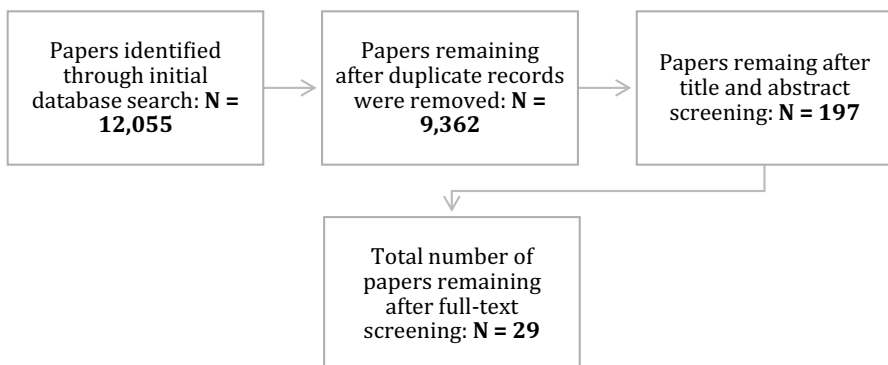


Fig. 1 Stage-by-stage selection process

Results

Quality of studies

The first domain examined for quality was the study design of the papers. There were 20 studies (69%) that featured an experimental design with stated random allocation of participants between control and experimental group. The review featured nine studies (31%) that were quasi-experimental in nature, meaning there was non-random allocation of participants into experimental groups.

Only one of the studies featured participants being studied at more than one institution, with most of the studies included ($N=28$) only sampling from a single establishment. All studies produced response rates of over 75%, which means they were given the highest score in that domain.

In terms of the type of data presented, all included studies featured an objective measure of learning outcomes such as test scores or completion times. No studies used self-assessment on the part of the participant to gauge learning outcomes.

The most pronounced weakness of the studies included in the review was the validity of the evaluation instrument used to assess learning outcomes. This domain pertained to the physical assessment instrumentation such as the quiz, test, or questionnaire that was given to the participant. Only six of the included studies (21%) reported the internal structure sufficiently through dimensionality, measurement invariance, or reliability using the criteria set down by Rios and Wells (2014). In addition, only 10 studies (34%) stated how the content was validated, with the majority ($N=19$) not reporting this information. Only three studies (Kozhevnikov et al. 2013; Makransky et al. 2017; Molina-Carmona et al. 2018) appropriately outlined both the internal structure and validity of evaluation content. The majority of studies ($N=16$) did not report either item.

Of the 29 studies in the current review, 26 scored full marks on the data analysis domain with both an appropriate and sufficiently complex analysis and reporting of the findings. Three studies scored lower than this due to reporting descriptive statistics only (Angulo and de Velasco 2013; Babu et al. 2018; Ray and Deb 2016).

Overall, the average quality score of a study in this systematic review was 12.7 with a range of 10.5–14.5 ($SD=1.0$). This was 1.8 points higher than the review carried out by Jensen and Konradsen (2018), which could in part be due to differences in study design criteria which was previously outlined. A full summary of the MERSQI scores for each study can be found in Table 2 in the Appendix.

Subject areas and learning domains

Table 1 provides a summary of all 29 articles that were included in the review. Studies were first categorised by the population that was sampled. Most I-VR studies took place in a higher education establishment (college or university) using undergraduate or postgraduate students ($N=25$). A smaller number of

Table 1 Details of included studies including domain, variables, and summary of main findings

Study (year)	Learning domain	HMD(s) used	Comparison variable(s)	Findings
Harrington et al. (2018)	Cognitive	Samsung Gear VR	2D video	No difference in information retention between both mediums
Yoganathan et al. (2018)	Procedural	Unspecified	2D video Face-to-face training	I-VR produced higher scores in knot tying skills. No difference in completion time
Makransky et al. (2017)	Cognitive	Samsung Gear VR	Desktop VR	Knowledge test score worse when using I-VR. No difference in transfer score
Greenwald et al. (2018)	Cognitive	HTC Vive	2D system	Study found no difference in test scores or number of moves between both mediums. One session was slower in I-VR compared to 2D
Lamb et al. (2018)	Cognitive	HTC Vive	Lecture Serious game Hands-on activity	I-VR produced a better test score than a traditional lecture. No difference between I-VR and educational game or hands-on activity
Webster (2016)	Cognitive	Sony HMZ	Lecture	I-VR produced significantly higher test scores than a lecture
Smith et al. (2018)	Cognitive and procedural	Oculus Rift	Desktop VR Written instructions	No difference in test score or completion time between I-VR and other methods
Ostrander et al. (2018)	Cognitive	HTC Vive	Instructor lesson	In 6 out of 7 lessons, I-VR was no more effective than a lesson delivered by an instructor
Angulo and de Velasco (2013)	Cognitive	Other	3D physical model	Higher appraisal of architectural spaces designed with I-VR than physical model
Stepan et al. (2017)	Cognitive	Oculus Rift	Online textbook	No difference in test score at pre-test, post-test, or 8-week follow-up
Sankaranarayanan et al. (2018)	Cognitive and procedural	Oculus Rift	No I-VR (written material only)	Using I-VR produced higher learning gains in procedural activity and transfer test. There was no difference in MCQ test scores
Fogarty et al. (2017)	Cognitive	Oculus Rift	No I-VR (lectures only)	At pre-intervention the participants performed significantly poorer than the control group. At post-intervention, this difference disappeared showing improvement in cognitive knowledge

Table 1 (continued)

Study (year)	Learning domain	HMD(s) used	Comparison variable(s)	Findings
Rupp et al. (2019)	Cognitive	Google Cardboard Oculus CV1 Oculus DK2	2D video (on phone screen)	Significantly better learning outcomes were demonstrated on test scores when I-VR was used compared to non-immersive condition
Bharathi and Tucker (2015)	Procedural	Oculus Rift	Desktop VR	Functional analysis task was carried out significantly faster in I-VR condition compared to D-VR
Johnston et al. (2018)	Cognitive	Oculus Rift	No I-VR (lectures only)	Those participants who underwent the I-VR experience performed better on final exam question than those who did not use I-VR
Allcoat and von Mühlenen (2018)	Cognitive	HTC Vive	Textbook 2D Video	Those who used I-VR performed better on a test than 2D control group. The I-VR group performed better on <i>remembering</i> questions than those in the textbook group, but no difference in <i>understanding</i> questions
Kozhevnikov et al. (2013)	Cognitive	nVisor SX60	Desktop VR	Performance in understanding relative motion concepts were significantly higher when participants used I-VR compared to D-VR
Parong and Mayer (2018)	Cognitive	HTC Vive	Slideshow	I-VR was demonstrated to cause significantly poorer results on a cell biology test than those who learned using a traditional PowerPoint
Olmos-Raya et al. (2018)	Cognitive	Samsung Gear VR	Desktop VR	Those participants who learned about geography using I-VR demonstrated significantly higher learning gains and medium term retention than those who used D-VR
Akbulut et al. (2018)	Cognitive	Mobile VR	Lecturers and lab sessions	Test scores were significantly higher in those who used the I-VR platform compared to those who learned through lectures and lab sessions alone
Ray and Deb (2016)	Cognitive	Google Cardboard	Slideshow Lectures	Over the course of 16 sessions, those who learned computer science through I-VR tended to score higher on test scores than those who did not

Table 1 (continued)

Study (year)	Learning domain	HMD(s) used	Comparison variable(s)	Findings
Moro et al. (2017)	Cognitive	Oculus Rift	Desktop VR	I-VR and D-VR were shown to be equally effective modes of learning on an anatomy skill test
Maresky et al. (2019)	Cognitive	Oculus Rift	Independent study	Those who used I-VR to learn about cardiac anatomy scored significantly higher on post-test than a control group who used independent study
Madden et al. (2018)	Cognitive	Oculus Rift	Desktop VR Ball-and-stick method	I-VR was no more effective than D-VR or a traditional model demonstration when learning about astronomy
Liou and Chang (2018)	Cognitive	HTC Vive	Traditional didactic teaching	I-VR was shown to produce significantly higher levels of learning in anatomy and chemistry tests than those who did not use the technology
Molina-Carmona et al. (2018)	Cognitive	Mobile VR	Desktop VR	I-VR was shown to improve the spatial ability of those participants who used it, compared to those that used D-VR
Babu et al. (2018)	Cognitive	HTC Vive	Desktop VR (tablet based)	Although I-VR and D-VR produced similar post-test scores, I-VR was shown to improve retention of knowledge learned
Alhalabi (2016)	Cognitive	Oculus Rift	Traditional methods	I-VR was shown to produce significantly higher attainment in engineering content than traditional educational methods. Furthermore, HMD based I-VR produced better results than a CAVE based system
Gutiérrez-Maldonado et al. (2015)	Affective and cognitive	Oculus Rift DK1	Stereoscopic desktop VR	Both immersive and non-immersive mediums produced similar results in psychiatric interview training

studies used high school pupils ($N=2$), or adult education students ($N=2$) such as those in vocational or work-based programmes.

Each of the included studies were then examined for the topic and subject area they pertained to. This was based upon the nature of the VR experience, participant pool, and intervention. In total, six main subject areas were identified. This included: medicine ($N=4$), science (biology, chemistry, and physics) ($N=13$), social science (human geography) ($N=1$), computer science ($N=2$), engineering and architecture ($N=7$), and safety education ($N=1$). One of the included studies (Molina-Carmona et al. 2018) did not neatly fit into one of the pre-defined categories as it utilised I-VR to teach abstract spatial concept abilities to multimedia engineering students. It was therefore categorised as ‘other’. Figure 2 shows the percentage of papers included by subject area.

In addition to the subject area, the learning outcomes were also categorised into three specific domains based upon the findings of previous systematic reviews, as well as the taxonomy of learning developed by Bloom et al. (1956). The first was *cognitive* which related to studies that intended to teach specific declarative information or knowledge. The second was *procedural* which intends to teach the user how to perform a specific task or learn psychomotor skills that pertain to a certain activity. Finally, the third learning outcome was *affective* skills which can be defined as a growth in areas relating to emotion and attitude. Most of the included studies ($N=24$) concentrated on the cognitive domain, with two studies focusing on purely procedural and psychomotor skills. The remaining studies were a blend of two domains with Sankaranarayanan et al. (2018) and Smith et al. (2018) examining both cognitive and procedural skills, and Gutiérrez-Maldonado et al. (2015) utilising both cognitive knowledge and affective awareness in psychiatric diagnosis training. Figure 3 shows the percentage of studies included by learning domain.

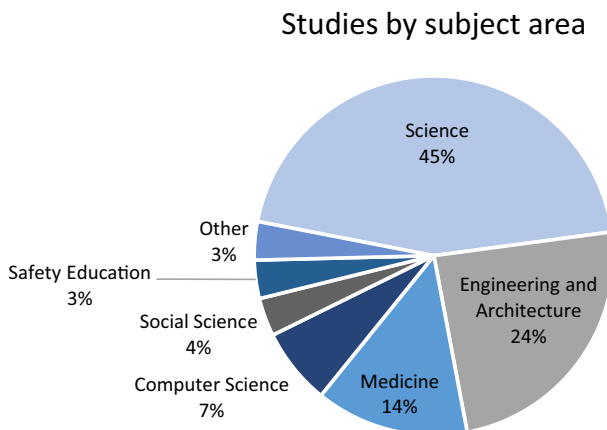


Fig. 2 Percentage of papers per subject area

Studies by learning domain

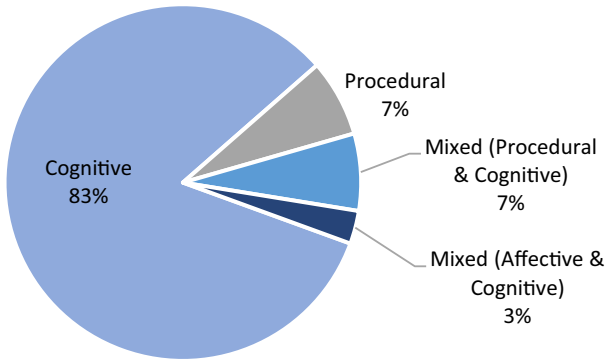


Fig. 3 Percentage of papers per learning domain

Experimental design

Outcome measures

A thorough understanding of the role of I-VR as a pedagogical practice can only be fully appreciated when consideration is given to the assessment instrumentation and outcome measures used to assess its utility. As previously mentioned, when analysing the quality of the included studies, it was the evaluation instrumentation itself that was shown to have the most pronounced weakness.

To assess the evaluation instruments being employed, the measures were broken down into two broad domains: outcome measures, and assessment instrumentation. Outcome measures can broadly be defined as how learning outcomes were quantified (e.g. by comparing test scores). Assessment instrumentation pertains to the evaluative instrument itself that is used to measure the learning outcomes (e.g. multiple-choice questionnaire, exam style questions). Twenty-seven of the included studies (93%) used test scores to assess learning outcomes, with the majority using this as their sole method. There were four studies that used completion time as a metric of learning outcome, although only one study (Bharathi and Tucker 2015) used this method exclusively. There was one study (Sankaranarayanan et al. 2018) that used the correct order of operation in a procedural task as one of its main outcome measures. There were three papers that utilised other outcome measures that could not be easily categorised. For instance Greenwald et al. (2018) used counting the number of moves needed to complete a task, Webster (2016) used the performance on a virtual jigsaw puzzle, and Angulo and de Velasco (2013) used a mixture of scores and evaluations of an architectural space.

Assessment instrumentation

In terms of the direct assessment instrumentation used to examine outcome measures, there was a heavy reliance on the multiple-choice questionnaire (MCQ). There were eighteen (62%) studies that utilised this method of assessment, with the majority of those using it as their sole evaluation instrument. Only five studies used extended answer questions (long or short form) to probe for a deeper understanding of the educational content, which was usually done in combination MCQs. The studies that included the teaching of procedural skills used marking criteria and checklists to assess whether the correct order was being followed. For instance Yogananathan et al. (2018) had an expert assessor use marking criteria to assess the knot tying skills of students. Similarly, Smith et al. (2018) had evaluators observe students with a decontamination checklist which evaluated performance based upon certain key tasks that were performed.

There were a smaller number of studies that used more novel instrumentation and methods for evaluation, such as the utilisation of labelling and identifying parts of a 3D model (e.g. Babu et al. 2018; Moro et al. 2017; Stepan et al. 2017). Fogarty et al. (2017) probed spatial and conceptual understanding in their assessment instrument by having participants draw shapes based on their understanding of structural engineering principles. Additionally, Alhalabi (2016) used quizzes on both mathematical knowledge, and the appropriate understanding of graphics and charts as an assessment measure for engineering students.

There were three studies (Liou and Chang 2018; Madden et al. 2018; Ray and Deb 2016) where the nature of the assessment instrumentation could not be definitively ascertained from the description.

The majority of studies (62%) utilised the pretest–posttest design by comparing the test scores pre-intervention with those after the I-VR experience. The remainder of the studies tended to assess post-intervention scores only, usually by comparing the difference in learning outcome between I-VR and one or more control group. Less conventional means of post-intervention comparison was sometimes utilised, such as Johnston et al. (2018) comparing the average score on a specific exam question that pertained to an I-VR experience that some student did or did not undertake.

There were four studies that examined the short to medium term retention rate of information and learning through follow-up assessment. This ranged from as soon as 1 day after the initial I-VR experience (Babu et al. 2018), through to 6 months post-intervention (Smith et al. 2018). Olmos-Raya et al. (2018) and Stepan et al. (2017) had follow-up assessments at 1-week and 8-weeks, respectively.

Intervention characteristics

In addition to having appropriate assessment measures, it is also important to examine the nature of the I-VR intervention itself. The most popular HMDs used were the Oculus ($N=13$) and HTC Vive ($N=7$). There were seven studies that used a form of mobile VR headset such as the Google Cardboard or Samsung Gear VR. In one study (Yogananathan et al. 2018) the exact HMD system used could not be

definitively ascertained. Figure 4 provides a breakdown of the HMDs used in the included studies.

Most studies (72%) featured only a single intervention with the I-VR experience, meaning that the student was exposed to the technology just once. There were a few exceptions to this, with Ostrander et al. (2018) having seven individual I-VR experiences in their manufacturing lesson, as well as Ray and Deb (2016) utilising smartphone based I-VR over the course of 16 sessions. Other studies allowed a greater degree of freedom in the number of interventions or times that a student could use I-VR. This was usually a result of time being dedicated to the technology through scheduled classes or lab times (e.g. Akbulut et al. 2018; Alhalabi 2016; Molina-Carmona et al. 2018). Despite this, the I-VR intervention was usually a single and isolated one.

As well as most of the studies featuring a single intervention, the exposure duration was also typically short, ranging from 6 to 30 mins. Generally, the exception to this was when the I-VR exposure lasted as long as it took the participant to complete a specific task, assessment, or procedure within the immersive environment (e.g. Babu et al. 2018; Bharathi and Tucker 2015; Greenwald et al. 2018; Sankaranarayanan et al. 2018). Molina-Carmona et al. (2018) supplemented the limited intervention duration by allowing participants to take the HMD away with them, so they could access the educational content for 2 weeks outside the classroom. However, just as with the number of interventions, exposure duration tended to be short, lasting on average 13 mins for those I-VR experiences that had a set time limit.

Most of the studies (62%) utilised I-VR as the sole method of learning, and did not combine the technology with additional pedagogical practices or materials to encourage learning. Only a limited number of studies (38%) supplemented the I-VR lesson by providing additional aids that were designed to complement the educational experience. For example, Smith et al. (2018) and Stepan et al. (2017) both had participants use web-based modules and textbooks in addition to the I-VR experience before testing them on learning outcomes. A number of the included studies

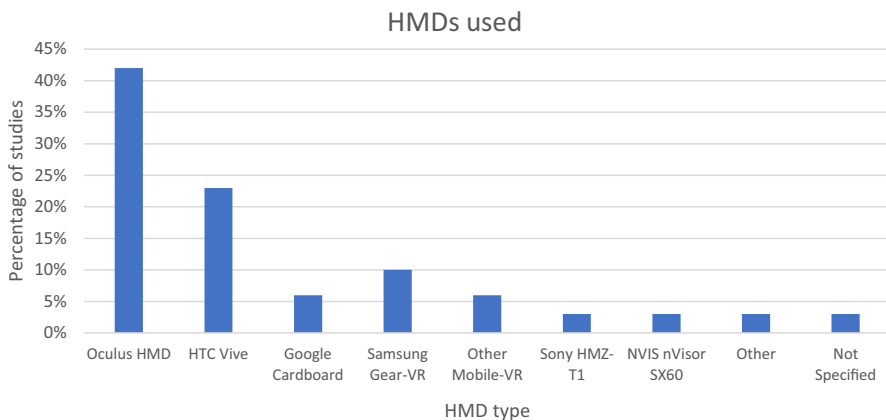


Fig. 4 HMDs used in studies

also utilised lecture based instruction or scheduled class time to operate in tandem with the I-VR environments (e.g. Akbulut et al. 2018; Fogarty et al. 2017; Johnston et al. 2018; Ray and Deb 2016; Sankaranarayanan et al. 2018).

Theoretical frameworks

A fundamental component of any educational tool or activity is to ground its use in learning theory or educational paradigms. Learning theories can broadly be broken down and defined by proposals regarding how student imbibe, process, and retain the information that they have learned (Pritchard 2017; Schunk 2011). When applied to educational I-VR, these theories should provide a pedagogical framework and foundation as how best to design interventions. Papers were examined for explicit statements regarding the theoretical basis for the study. Those papers that only mentioned theoretical approaches as part of the introduction or literature review were not deemed to have explicitly stated them. The majority of studies ($N=24$) made no mention of a theoretical approach underpinning the intervention. There were two studies that applied a *generative learning* framework (Makransky et al. 2017; Parong and Mayer 2018). This can be defined as an approach where the learner will actively integrate new knowledge with information that is already stored in the brain (Osborne and Wittrock 1985). Webster (2016) employed Mayer's (2009, 2014) *Cognitive Theory of Multimedia Learning* (CTML). CTML proposes a dual channel approach where visual and auditory information is actively processed, organised, and then stored in the brain. This is contingent on neither channel (visual or auditory) becoming overloaded with information. Smith et al. (2018) used the *NLN Jeffries Simulation Theory* as their theoretical basis. This theory, most commonly employed in nursing education, is where students learn information as part of a simulated experience (Jeffries et al. 2015). For the teaching of vocational skills, Babu et al. (2018) stated that their approach aligned with *situated learning*. Situated learning employs a constructivist approach in that students learns professional skills by actively participating in the experience (Huang et al. 2010).

Learning outcomes

For I-VR to gain wide-spread acceptance as a reliable pedagogical method, it must be shown to confer a tangible benefit in terms of learning outcomes over less immersive or traditional teaching methods.

Cognitive studies

There were twenty-four included studies that fell into the *cognitive* domain and aimed to teach specific declarative information or knowledge through the I-VR environment. The current review found that most studies demonstrated benefits in terms of learning outcomes when using I-VR compared to less immersive methods of learning. A smaller number of studies found no significant advantage regardless

of the pedagogical method being utilised. The results of these cognitive studies have been broken down by subject area.

Science based cognitive studies

The review found that cognitive learning activities requiring a high degree of visualisation and experiential understanding may be best facilitated using immersive technologies. For instance, both Liou and Chang (2018) and Maresky et al. (2019) found that anatomical learning facilitated by complex 3D visualisations of the human body were more conducive to learning in I-VR compared to traditional learning or independent study. Similarly Lamb et al. (2018) used a virtual environment that allowed for the manipulation and movement of strands of DNA, which produced better learning outcomes in content tests than a lecture or a serious educational game. Greater attention and engagement with the I-VR environment as measured with infrared spectroscopy was one of the possible explanations given for the effectiveness of the technology. In a study by Johnston et al. (2018), participants volunteered to take part in a cell biology experience either because they were engaged with the subject matter itself, or wanted supplementary instruction. Johnston et al. (2018) compared the exam scores of those students who volunteered to take part with those who did not. The study found that participants who underwent the I-VR experience scored 5% higher on the related exam question compared to the rest of the assessment. Those who did not undergo the cell biology I-VR experienced scored on average 35% worse on the same question.

The increase in graphical fidelity afforded by I-VR has allowed not only for the creation of complex computer-generated environments, but also the viewing of high resolution 360° video. In one such study, Rupp et al. (2019) had participants watch a six minute 360° video about the International Space Station with either a HMD which created a sense of immersion and presence, or on a mobile screen. The research found that those participants in the HMD condition scored significantly higher in a learning outcome test (MCQ) than those who watched the video in the non-immersive condition.

Although I-VR has been shown to confer a benefit in science education, there is evidence to suggest that not all learning objectives can be learned equally well. For instance, in task devised by Allcoat and von Mühlenen (2018), the researchers found that I-VR conferred a benefit over video or textbook learning when questions required *remembering*, but not ones pertaining to *understanding* of the material. The authors suggest that unfamiliarity and the novelty of the I-VR environments could have contributed to the lack of an obvious benefit in the latter domain. Another study that examined specific question types to understand I-VR's effectiveness was undertaken by Kozhevnikov et al. (2013). In this study, participants learned more conceptual and abstract relative motion concepts using either I-VR or D-VR. The study demonstrated that those in the I-VR condition performed significantly better in the two-dimensional problems than their D-VR counterparts, although there was no significant difference between groups in problems featuring only one spatial dimension.

There were several studies in the domain of science that showed no obvious benefits to using I-VR over traditional pedagogical methods. Two studies (Greenwald

et al. 2018; Moro et al. 2017) compared science learning in I-VR with desktop based VR and 2D videos. Results showed no clear benefit of I-VR based instruction when comparing the difference and significance of learning outcomes between mediums. Similarly, Stepan et al. (2017) found that I-VR was no more effective than online textbooks for the teaching of neuroanatomy. Interestingly, the same study found no difference in information retention rates when the participants were reassessed 8-weeks later. Madden et al. (2018) used I-VR, D-VR, and the traditional ball and stick method to teach astronomy principles pertaining to phases of the moon. The study found that I-VR and D-VR produced comparable test score results, with no significant differences in attainment. However, the authors commented on the encouraging finding that despite being a novel technology to most participants, I-VR still facilitated comparable learning outcomes to more traditional methods.

Despite the majority of studies demonstrating that I-VR learning is more effective or at least on par with traditional pedagogical methods, some studies have shown a detrimental effect of I-VR. Makransky et al. (2017) used a combination of assessment and EEG to find that an I-VR lab simulation produced significantly poorer test scores than a non-immersive alternative. Similarly, during another science experiment, Parong and Mayer (2018) found that students who used I-VR during a biology lesson scored significantly poorer than those who learned using a PowerPoint. Both of these studies cited Mayer's (2009, 2014) Cognitive Theory of Multimedia Learning as a possible explanation for the poorer performance for I-VR. The researchers postulate that the high-fidelity graphics and animations could have significantly increased cognitive load, which would have detracted from the learning task at hand. It was therefore proposed that a less immersive, yet well designed PowerPoint presentation would facilitate better learning outcomes than a graphically rich I-VR experience.

Engineering and architectural based cognitive studies

I-VR was effective in engineering and architectural education as a tool to visualise key concepts within the discipline. For example, Fogarty et al. (2017) allowed students to volunteer for an I-VR experience who struggled with the comprehension of spatial arrangements in structural engineering. Before the intervention, those students who volunteered to take part scored significantly poorer than their non-intervention counterparts. At post-test, not only did those who underwent the I-VR experience score significantly higher than they did at pre-test, but they also eliminated the significant difference with the non-intervention group. This would suggest that I-VR could serve an important function in supplementing or assisting learning in those students who are struggling to grasp complex problems relating to their discipline. Interestingly, Angulo and de Velasco (2013) used many of these same spatial and visualisation principles in a more applied setting. Their study split students into groups who were tasked with designing an architectural space (a health clinic waiting room), either with the assistance of an I-VR design tool (experimental group) or a physical model (control group). The study found the space that gained the most positive affect was designed by the I-VR group.

Webster (2016) created a graphically rich immersive environment which combined active and passive media with elements of gamification and interactivity to teach corrosion concepts to US army personnel. The study found that although both the I-VR environment and a traditional lecture were effective pedagogical methods for teaching these principles, it was the I-VR condition that produced the highest gain in knowledge acquisition.

There was also some evidence to suggest that I-VR interventions could assist in short-term retention of information in engineering related activities. Babu et al. (2018) found that although participants performed similarly in a mechanical labelling task using either I-VR or a tablet computer immediately post-intervention, the I-VR group had better retention of knowledge when the test was re-administered 1 day later. Furthermore, those participants in the I-VR group were also less likely to wrongly recall information compared to the non-immersive group on the retention test.

Interestingly, Ostrander et al. (2018) examined cognitive learning outcomes over seven separate manufacturing tasks utilising I-VR in one group, and a traditional class-based environment in the other. The study found that in six out of seven tasks, I-VR was no more effective than a traditional class where students could interact with the instructor or the physical models that they were accustomed to.

Medical based cognitive studies

Although papers featuring surgical simulators did not form part of this review, there were several applications of I-VR in the field of general medical education. Harrington et al. (2018) had medical students watch a ten-minute 360° video with slides containing surgical information superimposed over it. This was viewed either on a large television screen, or through a Gear VR headset. The study found no significant differences in knowledge retention scores between those who viewed the information through a HMD, or a traditional television screen. Despite not showing a distinct advantage in cognitive learning outcomes, the authors did suggest that the 360° surgical experience may facilitate a better understanding of how teamwork and interaction takes place within an operating theatre. This type of learning may be more difficult to measure using assessment instrumentation such as the MCQ, but nevertheless it could be that the experiential nature of I-VR may facilitate an understanding of interactions and communications. Smith et al. (2018) used either I-VR or D-VR on a computer to teach students about decontamination protocols. The research found that I-VR was no more effective than D-VR in a MCQ immediately post-intervention, or at 6-weeks follow-up.

Computer science based cognitive studies

Two studies demonstrated a significant advantage in using I-VR to teach computer science information. For instance, Akbulut et al. (2018) found that students who underwent an I-VR experience that focused on software engineering principles scored 12% higher than students who did not undergo I-VR learning. Interestingly, in a study by Ray and Deb (2016) that ran over 16 sessions on microcontrollers in

computing, the I-VR group performance lagged behind that of the control group who used slideshows for the first four sessions. It was only on session number five that the I-VR group outperformed the control group, and this performance enhancement remained relatively stable in the majority of the remaining 11 sessions. In effect, it took the I-VR group some time to *catch up* with the control group, but once they did, they tended to outperform them in the remaining lessons. The authors propose that this may have been due to the novelty of the I-VR equipment which participants took time to become comfortable and competent with.

Other cognitive studies

I-VR was also used by Molina-Carmona et al. (2018) as a means of spatial ability acquisition and visualisation. The study showed that learning outcomes as assessed by a spatial visualisation test were higher among those who undertook the task in an immersive, compared to a non-immersive environment. There was only one study in the field of social science that used I-VR to teach cognitive information. Olmos-Raya et al. (2018) used either I-VR or a tablet-based system to teach high school students about human geography. The research found that I-VR produced higher learning gains on a MCQ than the tablet-based system. Further, those who used I-VR performed better than the non-immersive group on a knowledge retention quiz when administered 1-week later.

Procedural studies

Three of the four studies that attempted to utilise I-VR as a means of teaching procedural skills showed a distinct advantage over less immersive methods. Bharathi and Tucker (2015) found that engineering students were faster in assembling a household appliance in a virtual functional analysis activity in I-VR compared to D-VR. Yoganathan et al. (2018) also found that medical students were more accurate in knot tying practice when using I-VR as a training tool as opposed to a control group who used a standard video. Medical and surgical residents were also studied by Sankaranarayanan et al. (2018) who used I-VR as a teaching tool for emergency fire response in an operating theatre environment. This study found that 70% of those who utilised the I-VR training were able to perform the correct procedure in the correct order. This was 50% higher than the control group who were exposed to a presentation and reading material only and did not experience I-VR.

One of the studies found no significant advantage to using I-VR as a learning tool. Smith et al. (2018) split nursing students into an I-VR group, a D-VR group (desktop PC based), or a written instruction group to learn about appropriate protocols for decontamination. The study found that there was no significant difference in performance between the groups as measured by a decontamination checklist, or the time taken to complete the task. Furthermore, reassessment 6 months later showed that I-VR conferred no advantage in procedural knowledge retention (accuracy and speed) compared to less immersive methods.

Affective studies

Only one of the studies attempted to use I-VR as a pedagogical tool to teach applied behavioural and affective skills. Gutiérrez-Maldonado et al. (2015) used I-VR in the field of diagnostic psychiatry in an attempt to improve interview skills when assessing patients for an eating disorder. Participants were exposed to a series of virtual patient avatars in either the I-VR condition, or a D-VR condition using stereoscopic glasses. Analysis showed that both conditions were equally as effective, and no significant differences were shown in the acquisition of skills between the two groups. Nevertheless, this was a novel study as it traversed the boundaries between traditional cognitive skill acquisition and applied behavioural and affective change.

Discussion and implications

The purpose of this review was to investigate I-VR's effectiveness as a pedagogical method in education, as well as examining the experimental design and characteristics of the included studies. In particular, the review found that the utilisation of I-VR is typically restricted to a small number of subject areas such as science and engineering. Furthermore, a heavy reliance has been placed on the MCQ and test score measures to assess learning outcomes. In addition, I-VR interventions were typically short and isolated, and were not complemented with additional or supplementary learning material. Despite this, most studies did find a significant advantage of using I-VR over less immersive methods of learning. This was the case particularly when the subject area was highly abstract or conceptual, or focused on procedural skills or tasks.

Is the utilisation of I-VR within education restrictive?

The findings of the review suggest a relatively homogenous application of I-VR in terms of both the subject areas represented, as well as the learning domain being taught. Almost 70% of the studies were from the field of science or engineering, with other subjects being marginally represented. It is worth noting, however, that although medical disciplines made up a small proportion of the studies included (14%), this was because most medical applications of I-VR feature surgical simulators and therefore were not part of the current review's inclusion criteria. Most studies utilised I-VR as a way of teaching cognitive skills, with only a handful examining the procedural or affective applications.

The findings of the review raise several issues when trying to assess the general effectiveness of I-VR in education. Similar to the findings of others (e.g. Jensen and Konradsen 2018; Radianti et al. 2020), the arts, humanities, and social sciences were underrepresented in the current review. This makes generalisable conclusions as to the cognitive benefit of the technology in these subjects challenging. One major reason for this under representation may be the lack of I-VR learning content, experiences, and teaching tools. Jensen and Konradsen (2018) highlighted that

instructors are restricted to the material published and produced by VR designers, and this may not necessarily meet the individual needs of the teacher, or the learning outcome trying to be achieved. The skillset needed to produce and create wholly virtual environments that can be rendered and displayed in a HMD is still demanding, despite the release of affordable VR creation suites. Therefore, the bespoke I-VR experiences required to teach social science lessons (or indeed any subject) is completely dependent on an appropriate I-VR tool already existing or having the technical proficiency to create one. A potential solution to the lack of bespoke material could be the examination of the pedagogical effectiveness of HMD 360° video in the classroom, as opposed to computer-generated environments. This content is comparatively easier to create using appropriate video equipment and can be tailored to the individual needs of the instructor or student group. Widespread research that examines the potential of I-VR in a multitude of diverse disciplines and learning domains will continue to be constrained by the availability of the requisite material. That is until such a time where bespoke and individually tailored I-VR experiences become more accessible.

Implications of outcome measures and assessment instrumentation

One of the most striking characteristics of the assessment instrumentation used in the studies was the reliance on the MCQ to assess learning outcomes. Although there have been many debates on the respective advantages and disadvantages of utilising the MCQ, it has generally been considered that it is most appropriate for testing large amounts of surface knowledge over the course of an entire module or syllabus (Excell 2000). As O'Dwyer (2012) points out, the assessment instrumentation encourages comprehensive learning of the entirety of the taught material, as opposed to just specific components. However, since most of the studies featured single interventions of between 6 and 30 mins, doubts are cast on whether MCQs are the most appropriate way to assess learning. Since the MCQ was most commonly administered immediately after the I-VR experience, much of the information may still be stored in short-term memory, and this may not give an accurate reflection of more comprehensive learning or long-term retention.

A second disadvantage associated with the heavy reliance on the MCQ is the limited breadth of knowledge that can be assessed. In Jensen and Konradsen's (2018) systematic review, the researchers found that none of the cognitive studies went beyond teaching *lower level* cognitive skills as defined by Bloom's taxonomy (Bloom et al. 1956). Similar results were found in the current review, with most studies requiring only a knowledge of previously learned material to successfully achieve the desired learning goal. Previous research on pedagogical assessment material (e.g. Ozuru et al. 2013) has suggested that the MCQ cannot assess higher levels of cognitive understanding or conceptual knowledge. Therefore, it may not only be the nature of the I-VR experience itself that restricts the learning of higher level cognitive skills, but also the restrictive nature of the assessment instrumentation that may impede an appropriate demonstration of learning outcomes. The utilisation of short or long form answers could be able to provide a more appropriate

measure of the depth of learning achieved, giving the student an opportunity to demonstrate their conceptual knowledge of a given subject. Furthermore, I-VR research could benefit by expanding the very definition of what constitutes a learning outcome. This could be achieved by not relying exclusively on test score comparisons, but rather examine how I-VR could be used to foster deeper conceptual understanding through experiential learning and subsequent classroom discussions with peers or instructors.

Implication of intervention characteristics for learning outcomes

The current review examined how I-VR is being utilised in experimental and applied settings, and the implications this has for assessing its pedagogical suitability. In most studies, the participant took part in a single I-VR experience that was also short in duration. This presents several key challenges. Most importantly, the novelty of the I-VR technology itself may have impeded the learning experience of the user, especially if they had never used the technology before or were unfamiliar with it. This seemed to be demonstrated by Ray and Deb (2016) who found that in the initial sessions of I-VR learning, performance was on average poorer than those who underwent traditional teaching methods. It was only after the participants began to become familiar with the technology (on session number five) that learning surpassed the control group. Similarly, studies that allowed for extended exposure to I-VR (e.g. Akbulut et al. 2018; Alhalabi 2016; Molina-Carmona et al. 2018), either through free navigation, repeated sessions, or scheduled class time, tended to show an advantage of using I-VR over non-immersive or traditional methods. It is therefore important to address the potentially negative influence that I-VR's novelty as a learning tool may have, especially when outcomes are directly compared to another medium or method. Scepticism for media comparison studies was highlighted in the 1980s by Clark (1983), and then later re-addressed by Parong and Mayer (2018). As Parong and Mayer (2018) put it, the side-by-side comparison of two learning methods is an "apples-to-oranges type of comparison" (p. 788). This "apples-to-oranges" comparison is made starker when considering that I-VR is an unfamiliar technology to most in an educational capacity, and its pedagogical outcomes are being directly compared with familiar methods such as textbooks or lectures. It is important to consider that the novelty of HMDs and I-VR may hinder learning outcomes and classroom application, and it is therefore prudent to ensure that the degree of familiarity with I-VR technology is factored into any direct comparison with other methods. In practice, this could mean that participants require extended familiarisation trials or free navigation before the start of experimental studies as a means of mitigating against potential problems caused by technological novelty.

In addition to the short intervention and exposure time, most studies did not complement I-VR with an additional method of teaching or self-learning. The limited number of studies that did tended to utilise web-based textbooks or modules, as well as lectures and scheduled class time. Encouragingly, those studies that combined or supplemented traditional class-based learning with I-VR (e.g. Akbulut et al. 2018; Fogarty et al. 2017; Johnston et al. 2018; Sankaranarayanan et al. 2018; Yoganathan

et al. 2018) tended to show a learning advantage. This suggests that I-VR may be best employed as form of blended or multi-modal learning to supplement and complement class-based instruction (Garrison and Kanuka 2004). An area for investigation would be to examine I-VR's application longitudinally in a natural classroom environment. The current review contained only a limited number of studies that employed this approach, however, by implementing and studying how I-VR can be adopted and integrated into a module or syllabus, a clearer picture of its capabilities can emerge.

Learning theories ultimately provide a theoretical framework and foundation as how best to design educational interventions (Pritchard 2017; Schunk 2011). However, the review found that few papers explicitly state that any predetermined learning theory was used to advise the characteristics or methods of the study. Similar findings were reported in a systematic review by Radianti et al. (2020) examining I-VR use in higher education exclusively. Radianti et al.'s (2020) review found that in around 70% of the 38 studies included, no learning theory was mentioned as forming the foundation of the VR activity. Several studies have shown that educators regard clear pre-defined intervention characteristics and objectives as essential components of I-VR teaching (Fransson et al. 2020; Lee and Shea 2020). It is therefore essential that future experimental and applied research is based on a sound theoretical basis that can advise how the technology can be appropriately utilised and assessed.

Learning outcomes in I-VR

The current review examined learning outcomes across three domains: cognitive, procedural, and affective. By far the most popular domain was the teaching of cognitive skills and knowledge which made up 83% of the studies in the current review. Around half of those demonstrated a positive effect on learning when using I-VR over less immersive pedagogical methods. Most of the remaining studies showed no significant effect either way, with only a small number of papers exhibiting detrimental results. Researchers have suggested that the increased levels of immersive content that stimulate multisensory engagement can ultimately lead to more effective learning outcomes (Webster 2016). When this is implemented in cognitive learning activities that require a high degree of spatial understanding and visualisation (e.g. Maresky et al. 2019), I-VR can allow the user to gain insights that are difficult to reproduce in reality. This review has already identified scientific subjects such as biology and physics as promising avenues for educational I-VR implementation. However, other scientific disciplines that require abstract or conceptual understanding (e.g. chemistry, mathematics) could also benefit from the visualisation afforded by I-VR.

Studies that utilised I-VR for the teaching of procedural skills and knowledge produced encouraging results, with three of the four studies finding a significantly positive increase in learning (Bharathi and Tucker 2015; Sankaranarayanan et al. 2018; Yoganathan et al. 2018). Interestingly, two of the studies featured a transfer component by having the user first practice the procedure in I-VR, and then use this

form of experiential learning to complete a task in the real world. Yoganathan et al. (2018) had students practice how to tie a surgical knot in I-VR and then complete this task for real in-front of an expert. Sankaranarayanan et al. (2018) had medical students learn how to deal with an operating theatre fire by first practicing the procedure in I-VR, and then applying this knowledge to a mock emergency in a real operating room. Both studies found a positive effect of using I-VR as the training method by demonstrating improved results when performed in a real environment. These are encouraging findings for I-VR's effectiveness in psychomotor and procedural education, as there has been a degree of scepticism over whether I-VR simply produces a "getting good at the game" effect. For instance, Jensen and Konradsen (2018) point out that the honing of procedural skills within I-VR may simply lead to the participant becoming proficient when performing the task virtually, and this may not necessarily transfer to the real world. The current review has identified that the two procedural studies that implemented a transfer task did indeed demonstrate a significant benefit to using I-VR as an initial education method. This demonstrates that virtual training can be a successful precursor to implementation in the real world. This suggests that I-VR could be useful in educating students in dangerous vocational subjects such as electrical engineering without risk to themselves or others. However, this view is based on a small number of studies, and it is therefore important that future procedural tasks utilise a transfer activity to understand the potential scope and parameters surrounding I-VR training and real-world application.

Only one of the studies had a firm focus on the training of affective skills, namely by using I-VR as a way of teaching diagnostic interview techniques in a psychiatric setting (Gutiérrez-Maldonado et al. 2015). Although this study found no clear advantage to using I-VR, other research out with the domain of education has demonstrated promising results in utilising the technology for affective and behavioural change. This included applying the technology successfully in areas such as exposure therapy, anxiety disorder treatment, and empathy elicitation (Botella et al. 2017; Maples-Keller et al. 2017a, b; Schutte and Stilinović 2017). As a result of the strong non-educational body of literature suggesting I-VR can facilitate affective and behavioural change, future research should examine how this can be applied in an educational context, and then transferred to real-world scenarios. For instance, in their psychiatric interview experience, Gutiérrez-Maldonado et al. (2015) had users interact solely with virtual avatars, and did not have the participants demonstrate their learning with a real actor or patient. Therefore, just like with procedural skill acquisition, affective I-VR experiences should seek to understand how virtual learning can then be applied to real situations.

Implications and future practice

The current review has been able to identify a body of experimental and applied research that show the potential benefits of using I-VR in education. It has already been noted that I-VR has traditionally been used to teach low level or fundamental skills and knowledge, and has not necessarily been used to facilitate what Bloom et al. (1956) would consider higher level learning. This would include analysing and

evaluating experience. By expanding the definition of learning outcomes to encompass potential benefits such as an increased depth of understanding or the ability to identify complex themes, pedagogical practice can take advantage of the inherent strength of the medium. These should be comprehensively analysed to investigate learning outcomes that go beyond simple test scores.

The review has also been able to identify areas for improvement in future studies, which would address confounding variables and expand the scope of research. Firstly, as Allcoat and von Mühlenen (2018) suggest, the novelty of I-VR could hamper learning outcomes due to unfamiliarity with the technology. Therefore, it is important to factor in an extended familiarisation or free navigation period that would help alleviate this concern. Additionally, follow-up qualitative analysis such as interviews or focus groups could help explore the phenomenology or direct experience of using I-VR, and highlight concerns relating to unfamiliarity or technological anxiety. The biggest concern relating to the assessment instrumentation was the over reliance on the MCQ (62% of studies used it as the sole method of assessment). Although this method is deemed appropriate for assessing large amounts of surface knowledge, it may not reveal more nuanced forms of learning that extend beyond mere recall of information. Therefore, long form essay questions, oral examinations, or group discussions could be used to facilitate students' ability to present their in-depth understanding and applied knowledge. Future research must base the nature of these interventions on a sound theoretical framework. This would assist in identifying specific learning objectives and methods of assessments. An explicit theoretical approach was commonly lacking in the included studies.

I-VR has already been demonstrated to be an effective tool in non-pedagogical behaviour change, such as treating phobias, mental health conditions, or as a tool for rehabilitation (Botella et al. 2017; Maples-Keller et al. 2018; Ravi et al. 2017). Research should therefore concentrate on I-VR's potential as an acquisition tool for affective skills. There is already a strong body of evidence suggesting I-VR experiences can elicit high levels of empathetic response and perspective taking, and this should be explored within an educational context (Herrera et al. 2018; Shin 2018). For example, Dyer et al. (2018) used I-VR to allow health care students to take the perspective of an older patient with age-related medical conditions, which led to increased empathy. Future studies should investigate whether this perspective taking ability can lead to higher domains of learning, such as evaluating one's actions, applying problem solving skills, or creating new solutions as a direct result of the insights they received from I-VR. This will require researchers and instructors to carefully consider their tools for evaluation and assessment, perhaps incorporating mixed-methods to give a more holistic overview of learning achieved.

Conclusions

The current review found that I-VR conferred a learning benefit in around half of cognitive studies, especially where highly complex or conceptual problems required spatial understanding and visualisation. Although many studies found no significant benefit of using I-VR over less immersive technology, only a small number resulted

in detrimental effects on learning outcomes. However, the homogenous nature of assessment instrumentation, such as an over reliance on the MCQ may have stifled the ability for participants to demonstrate learning outcomes beyond low level cognitive knowledge. Short exposure times and isolated interventions could also pose a problem as the novel nature of the technology could negatively impact the amount of learning able to be imbibed. Encouragingly, most procedural tasks did show a benefit to utilising I-VR, and furthermore, there was evidence that virtual skill acquisition could be transferred successfully to real world problems and scenarios. The ability to repeatedly practice a procedure in a safe environment whilst expending little resources could be one of the most advantageous and intrinsic benefits of I-VR technology. Although affective behavioural change has been widely studied in non-educational applications of I-VR, the domain was underrepresented in the current review, and is an important area for future investigation.

Over the coming years, technological advancement, an increase in creative content, and the possibilities for instructors to create bespoke I-VR experiences will all contribute to I-VR's potential as a teaching tool. It is essential therefore that the implementation of such technology is based on sound theoretical and experimental evidence in order to ensure that the I-VR is utilised correctly, and to its full potential.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Table 2.

Table 2 MERSQI quality appraisal scores for included studies

Study (year)	Design	Sampling	Type of data	Validity of instrument	Data analysis	Outcomes	Total
Harrington et al. (2018)	3	2	3	1	3	1.5	13.5
Yoganathan et al. (2018)	3	2.5	3	1	3	1.5	14
Makransky et al. (2017)	3	2	3	2	3	1.5	14.5
Greenwald et al. (2018)	2	2	3	0	3	1.5	11.5
Lamb et al. (2018)	3	2	3	0	3	1.5	12.5
Webster (2016)	2	2	3	1	3	1.5	12.5
Smith et al. (2018)	2	2	3	1	3	1.5	12.5
Ostrander et al. (2018)	3	2	3	0	3	1.5	12.5
Angulo and de Velasco (2013)	2	2	3	0	2	1.5	10.5
Stepan et al. (2017)	3	2	3	1	3	1.5	13.5
Sankaranarayanan et al. (2018)	2	2	3	1	3	1.5	12.5
Fogarty et al. (2017)	2	2	3	1	3	1.5	12.5
Rupp et al. (2019)	3	2	3	0	3	1.5	12.5
Bharathi and Tucker (2015)	3	2	3	0	3	1.5	12.5
Johnston et al. (2018)	2	2	3	0	3	1.5	11.5
Allcoat and von Mühlengen (2018)	3	2	3	1	3	1.5	13.5
Kozhevnikov et al. (2013)	3	2	3	2	3	1.5	14.5
Parong and Mayer (2018)	3	2	3	1	3	1.5	13.5
Olmos-Raya et al. (2018)	3	2	3	0	3	1.5	12.5
Akbulut et al. (2018)	2	2	3	0	3	1.5	11.5
Ray and Deb (2016)	3	2	3	0	2	1.5	11.5
Moro et al. (2017)	3	2	3	1	3	1.5	13.5
Maresky et al. (2019)	3	2	3	0	3	1.5	12.5
Madden et al. (2018)	3	2	3	0	3	1.5	12.5
Liou and Chang (2018)	3	2	3	0	3	1.5	12.5
Molina-Carmona et al. (2018)	3	2	3	2	3	1.5	14.5
Babu et al. (2018)	3	2	3	0	2	1.5	11.5
Alhalabi (2016)	2	2	3	0	3	1.5	11.5
Gutiérrez-Maldonado et al. (2015)	3	2	3	0	3	1.5	12.5
Average	2.7	2.0	3.0	0.6	2.9	1.5	12.7

References

- Akbulut, A., Catal, C., & Yıldız, B. (2018). On the effectiveness of virtual reality in the education of software engineering. *Computer Applications in Engineering Education*, 26, 918–927.
- Alaraj, A., Lemole, M., Finkle, J., Yudkowsky, R., Wallace, A., Luciano, C., ... Charbel, F. (2011). Virtual reality training in neurosurgery: Review of current status and future applications. *Surgical Neurology International*, 2, 52.
- Alhalabi, W. S. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, 35, 919–925.
- Allcoat, D., & von Mühlén, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26, 2140.
- Angulo, A., & de Velasco, G. V. (2013). Immersive simulation of architectural spatial experiences. *Blucher Design Proceedings*, 1, 495–499.
- Babu, S., Krishna, S., Unnikrishnan, R., & Bhavani, R. (2018). Virtual reality learning environments for vocational education: A comparison study with conventional instructional media on knowledge retention. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 385–389). IEEE.
- Bailenson, J. N., Markowitz, D. M., Pea, R. D., Perone, B. P., & Laha, R. (2018). Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in Psychology*, 9, 2364.
- Bharathi, A. K. B. G., & Tucker, C. S. (2015). Investigating the impact of interactive immersive virtual reality environments in enhancing task performance in online engineering design activities. In *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. ASME.
- Bloom, B. S., Englehard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York, NY: Longmans.
- Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: A systematic review. *Current Psychiatry Reports*, 19, 42.
- Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *TechTrends*, 60, 517–519.
- Çalışkan, O. (2011). Virtual field trips in education of earth and environmental sciences. *Procedia-Social and Behavioral Sciences*, 15, 3239–3243.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53, 445–459.
- Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: Two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology*, 36, 439–451.
- Dyer, E., Swartzlander, B. J., & Gugliucci, M. R. (2018). Using virtual reality in medical education to teach empathy. *Journal of the Medical Library Association*, 106, 498–500.
- Ellinger, R. S., & Frankland, P. (1976). Computer-assisted and lecture instruction: A comparative experiment. *Journal of Geography*, 75, 109–120.
- Excell, P. S. (2000). Experiments in the use of multiple-choice examinations for electromagnetics-related topics. *IEEE Transactions on Education*, 43, 250–256.
- Fogarty, J., McCormick, J., & El-Tawil, S. (2017). Improving student understanding of complex spatial arrangements with virtual reality. *Journal of Professional Issues in Engineering Education and Practice*, 144, 04017013.
- Fransson, G., Holmberg, J., & Westelius, C. (2020). The challenges of using head mounted virtual reality in K-12 schools from a teacher perspective. *Education and Information Technologies*, 2, 20–22.
- Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. In *The 11th International Scientific Conference ELearning and Software for Education* (pp. 133–141).
- Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *Internet and Higher Education*, 7, 95–105.
- Gee, J. P. (2004). What video games have to teach us about learning and literacy. *Education Training*, 46, 20.

- Greenwald, S. W., Corning, W., Funk, M., & Maes, P. (2018). Comparing learning in virtual reality with learning on a 2D screen using electrostatics activities. *Journal of Universal Computer Science*, 24, 220–245.
- Gutiérrez-Maldonado, J., Ferrer-García, M., Pla-Sanjuanelo, J., Andrés-Pueyo, A., & Talarn-Caparrós, A. (2015). Virtual reality to train diagnostic skills in eating disorders. Comparison of two low cost systems. *Studies in Health Technology and Informatics*, 219, 75–81.
- Häfner, P., Dückler, J., Schlatt, C., & Ovtcharova, J. (2018). Decision support method for using virtual reality in education based on a cost-benefit-analysis. In *4th International Conference of the Virtual and Augmented Reality in Education, VARE 2018*, 103–112.
- Harrington, C. M., Kavanagh, D. O., Wright Ballester, G., Wright Ballester, A., Dicker, P., Traynor, O., ... Tierney, S. (2018). 360° operative videos: A randomised cross-over study evaluating attentiveness and information retention. *Journal of Surgical Education*, 75, 993–1000.
- Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zak, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE*, 13, e0204494.
- Hodgson, E., Bachmann, E. R., Vincent, D., Zmuda, M., Waller, D., & Calusdian, J. (2015). WeaVR: A self-contained and wearable immersive virtual environment simulation system. *Behavior Research Methods*, 47(1), 296–307.
- Hodgson, P., Lee, V. W. Y., Chan, J. C. S., Fong, A., Tang, C. S. Y., Chan, L., et al. (2019). Immersive virtual reality (IVR) in higher education: Development and implementation. In M. Dieck & T. Jung (Eds.), *Augmented reality and virtual reality* (pp. 161–173). New York: Springer.
- Howard-Jones, P., Ott, M., van Leeuwen, T., & De Smedt, B. (2015). The potential relevance of cognitive neuroscience for the development and use of technology-enhanced learning. *Learning, Media and Technology*, 40, 131–151.
- Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers and Education*, 55, 1171–1182.
- Jeffries, P. R., Rodgers, B., & Adamson, K. (2015). NLN Jeffries simulation theory: Brief narrative description. *Nursing Education Perspectives*, 36, 292–293.
- Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23, 1515–1529.
- Johnston, A. P. R., Rae, J., Ariotti, N., Bailey, B., Lilja, A., Webb, R. I., ... Parton, R. G. (2018). Journey to the centre of the cell: Virtual reality immersion into scientific data. *Traffic*, 19, 105–110.
- Kozhevnikov, M., Gurliitt, J., & Kozhevnikov, M. (2013). Learning relative motion concepts in immersive and non-immersive virtual environments. *Journal of Science Education and Technology*, 22, 952–962.
- Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers and Education*, 124, 14–26.
- Larsen, C. R., Oestergaard, J., Ottesen, B. S., & Soerensen, J. L. (2012). The efficacy of virtual reality simulation training in laparoscopy: A systematic review of randomized trials. *Acta Obstetrica et Gynecologica Scandinavica*, 91, 1015–1028.
- Lee, C., & Shea, M. (2020). Exploring the use of virtual reality by pre-service elementary teachers for teaching science in the elementary classroom. *Journal of Research on Technology in Education*, 52, 163–177.
- Lee, E. A. L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, 55, 1424–1442.
- Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media*, 26, 71–86.
- Li, L., Yu, F., Shi, D., Shi, J., Tian, Z., Yang, J., ... Jiang, Q. (2017). Application of virtual reality technology in clinical medicine. *American Journal of Translational Research*, 9, 3867–3880.
- Liou, W., & Chang, C. (2018). Virtual reality classroom applied to science education. In *2018 23rd International Scientific-Professional Conference on Information Technology (IT)* (pp. 1–4).
- Lyne, D. S. (2013). Development of virtual reality applications for the construction industry using the Oculus Rift head mounted display. In *Proceedings of the 13th International Conference on Construction Applications of Virtual Reality* (pp. 30–31).
- Madden, J. H., Won, A. S., Schuldt, J. P., Kim, B., Pandita, S., Sun, Y., ... Holmes, N. G. (2018). Virtual reality as a teaching tool for moon phases and beyond. In *2018 Physics Education Research Conference*.

- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2017). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236.
- Maples-Keller, J. L., Bunnell, B. E., Kim, S.-J., & Barbara, R. (2017a). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25, 103–113.
- Maples-Keller, J. L., Yasinski, C., Manjin, N., & Rothbaum, B. O. (2017b). Virtual reality-enhanced extinction of phobias and post-traumatic stress. *Neurotherapeutics*, 14, 554–563.
- Maresky, H. S., Oikonomou, A., Ali, I., Ditkofsky, N., Pakkal, M., & Ballyk, B. (2019). Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. *Clinical Anatomy*, 32, 238–243.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education*, 70, 29–40.
- Molina-Carmona, R., Pertegal-Felices, M., Jimeno-Morenilla, A., & Mora-Mora, H. (2018). Virtual reality learning activities for multimedia students to enhance spatial ability. *Sustainability*, 10, 1074.
- Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, 10, 549–559.
- Murcia-López, M., & Steed, A. (2016). The effect of environmental features, self-avatar, and immersion on object location memory in virtual environments. *Frontiers in ICT*, 3, 1–10.
- O'Dwyer, A. (2012). A teaching practice review of the use of multiple-choice questions for formative and summative assessment of student work on advanced undergraduate and postgraduate modules in engineering. *The All Ireland Journal of Teaching and Learning in Higher Education*, 4, 1–12.
- Olmos, E., Cavalcanti, J. F., Soler, J.-L., Contero, M., & Alcañiz, M. (2018). Mobile virtual reality: A promising technology to change the way we learn and teach. In S. M. Bell (Ed.), *Mobile and ubiquitous learning: An international handbook* (pp. 95–106). Singapore: Springer.
- Olmos-Raya, E., Ferreira-Cavalcanti, J., Contero, M., Castellanos, M. C., Giglioli, I. A. C., & Alcañiz, M. (2018). Mobile virtual reality as an educational platform: A pilot study on the impact of immersion and positive emotion induction in the learning process. *Eurasia Journal of Mathematics, Science and Technology Education*, 14, 2045–2057.
- Osborne, R., & Wittrock, M. (1985). The generative learning model and its implications for science education. *Studies in Science Education*, 12, 59–87.
- Ostrander, J. K., Tucker, C. S., Simpson, T. W., & Meisel, N. A. (2018). Evaluating the effectiveness of virtual reality as an interactive educational resource for additive manufacturing. In *Volume 3: 20th International Conference on Advanced Vehicle Technologies; 15th International Conference on Design Education*, V003T04A018. ASME.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67, 215–227.
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110, 785–797.
- Pritchard, A. (2017). *Ways of learning: Learning theories for the classroom* (4th ed.). London: Routledge.
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education*, 147, 103778.
- Ravi, D. K., Kumar, N., & Singhi, P. (2017). Effectiveness of virtual reality rehabilitation for children and adolescents with cerebral palsy: An updated evidence-based systematic review. *Physiotherapy*, 103, 245–258.
- Ray, A. B., & Deb, S. (2016). Smartphone based virtual reality systems in classroom teaching - A study on the effects of learning outcome. In *2016 IEEE Eighth International Conference on Technology for Education (T4E)* (pp. 68–71).
- Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA*, 298, 1002.
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26, 108–116.

- Rupp, M. A., Odette, K. L., Kozachuk, J., Michaelis, J. R., Smither, J. A., & McConnell, D. S. (2019). Investigating learning outcomes and subjective experiences in 360-degree videos. *Computers & Education*, 128, 256–268.
- Sankaranarayanan, G., Wooley, L., Hogg, D., Dorozhkin, D., Olasky, J., Chauhan, S., ... Jones, D. B. (2018). Immersive virtual reality-based training improves response in a simulated operating room fire scenario. *Surgical Endoscopy and Other Interventional Techniques*, 32, 3439–3449.
- Schunk, D. H. (2011). *Learning theories: An educational perspective* (6th ed.). London: Pearson.
- Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, 41, 708–712.
- Sherman, W. R., & Craig, A. B. (2003). Understanding virtual reality: Interface, application, and design. In *Understanding Virtual Reality: Interface, Application, and Design*.
- Shin, D. (2018). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior*, 78, 64–73.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109, 431–433.
- Smith, S. J., Farra, S. L., Ulrich, D. L., Hodgson, E., Nicely, S., & Mickle, A. (2018). Effectiveness of two varying levels of virtual reality simulation. *Nursing Education Perspectives*, 39, 10–15.
- Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., et al. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International Forum of Allergy and Rhinology*, 7, 1006–1013.
- UKAuthority. (2019). VR and AR attract education sector interest. Retrieved December 16, 2019, from <https://www.ukauthority.com/articles/vr-and-ar-attract-education-sector-interest/>
- Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24, 1319–1333.
- Yoganathan, S., Finch, D. A., Parkin, E., & Pollard, J. (2018). 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *International Journal of Surgery*, 54, 24–27.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

D. Hamilton is a PhD student in psychology at the University of the West of Scotland. His current research focuses on the application of immersive virtual-reality technology as a pedagogical method in higher education.

J. McKechnie is a Professor of Psychology at the University of the West of Scotland. His research focuses on national and international aspects of child workers and early experiences of the workplace. As a founder member of the Child Employment Research Group, his work has included government funded projects into the nature and extent and policy and practice relating to child employment

E. Edgerton is a Reader in Psychology at the University of the West of Scotland. He is an experienced researcher in environmental psychology and has played a leading role in a wide range of environment-behaviour projects including: environmental design within a psychiatric hospital, the role of the learning environment in the educational process and dementia friendly design in care homes. He was a Board member of the International Association of People-environment Studies (IAPS) from 2005 to 2014 and was President from 2012 until 2014

C. Wilson is a Lecturer in Psychology at the University of the West of Scotland. Her research interests lie predominantly within the field of social and educational psychology. Claire is interested in applying psychological theories most commonly applied in social psychology to educational problems. She has examined attitudinal and social cognitive predictors of behaviour, automaticity of behaviour, behaviour change interventions and has used cognitive tests to assess psychological constructs across a range of contexts including teachers' inclusive behaviours for children with disabilities, bullying in schools and parent-child interactions